Research Article Open Access

Keywords: Genome-wide scan; Pathway; Association studies; Quantitative traits; HapMap project; Software

Abbreviations: GWAS: Genome-wide association studies; FDR: false discovery rate; CEU: Northern and Western Europe; YRI: Yoruba from Ibadan; CHB: Han Chinese from Beijing of China; JPT: Japanese from Tokyo of Japan

Introduction

Genome-wide association studies (GWAS) are very popular and successful for identifying disease genes by examining the relationship between each SNP and target traits in recent years [1]. In spite of the greater power of GWAS compared to linkage analysis, GWAS may miss the disease genes with weak genetic main effects or strong epistatic effects due to single locus testing approach [2]. To overcome these limitations, researchers developed pathway-based association study approaches, which combined the information of polymorphism and function of multiple related genes [3]. Pathway-based GWAS should be powerful for genetic studies of complex diseases, the susceptibility of which are determined by biological pathways.

Various pathway-based association study approaches have been proposed [3-5]. Evaluating the performance of different pathway-based association study approaches under certain parameter setting, such as sample sizes and disease genetic models, can provide guidelines for researchers to choose proper study methods and interpret their study results. Because the disease genes and genetic models of real population data are mostly uncertain or unknown in practice, simulations play an important role in the development of novel study approaches. However, to the best of our knowledge, there is few available simulating tool for pathway-based GWAS now.

In this study, we developed a flexible simulating tool PATHSIMU for pathway-based GWAS now. It can simultaneously simulate multiple quantitative phenotypes and genome-wide genotype data based on the real data from the HapMap project [6] or users.

Main Features

Genotype data

Real genetic data from the Hapmap project [6] or real GWAS studies can be used by PATHSIMU. The current version of PATHSIMU provides three choices for users to generate GWAS genotype data:

(1) We developed a genotype simulating program in PATHSIMU. The genome-wide SNP map, alleles and allele frequencies data of Utah residents with ancestry from Northern and Western Europe (CEU), Yoruba from Ibadan (YRI) of Africa, Han Chinese from Beijing of China (CHB) and Japanese from Tokyo of Japan (JPT) were downloaded from the HapMap website (http://hapmap.ncbi.nlm.nih.gov). Using the real SNP map, alleles and allele frequencies data of Hapmap, PATHSIMU can randomly simulate genotype for each SNP locus, and generate genome-wide genotype data with PLINK file format for CEU, YRI, CHB and JPT populations [7].

(2) Users' real GWAS data of complex diseases with PLINK file format [7] is also acceptable in PATHSIMU for following quantitative phenotype simulations.

(3) HAPGEN is another popular simulating tool [8-10], which can simulate genome-wide genotype data using the phased haplotype data, minor allele frequencies and linkage disequilibrium data of

***Corresponding author:** Xiong Guo, Key Laboratory of Environment and Gene Related Diseases of Ministry Education, College of Medicine, Xi'an Jiaotong

Citation: Zhang F, Guo X, Ma J (2012) PATHSIMU: A Flexible Simulating Tool for Pathway-based Genome-wide Association Studies. 1: 116. doi:10.4172/scientifcreports.116

Ha Ma : http://hapmap.ncbi.nlm.nih.gov/